Getting the most out of it: optimal experiments for parameter estimation of microalgae growth models

Rafael Munoz Tamayo, Pierre Martinon, Gaël Bougaran, Francis Mairet, Olivier Bernard

Getting the most out of it: optimal experiments for parameter estimation of microalgae growth models

Rafael Muñoz-Tamayo[a,], Pierre Martinon[b], Gaël Bougaran[c], Francis Mairet[a], Olivier Bernard[a,d]

[a]*BIOCORE-INRIA, BP93, 06902 Sophia-Antipolis Cedex, France.*
[b]*Commands INRIA Saclay, Ecole Polytechnique, CMAP 91128 Palaiseau, France.*
[c]*Ifremer, Laboratoire Physiologie et Biotechnologie des Algues, rue de l'île d'Yeu BP 21105 44311 Nantes cedex 3, France*
[d]*LOV-UPMC-CNRS, UMR 7093, Station Zoologique, B.P. 28 06234, Villefranche-sur-mer, France*

Abstract

Mathematical models are expected to play a pivotal role for driving microalgal production towards a profitable process of renewable energy generation. To render models of microalgae growth useful tools for prediction and process optimization, reliable parameters need to be provided. This reliability implies a careful design of experiments that can be exploited for parameter estimation. In this paper, we provide guidelines for the design of experiments with high informative content based on optimal experiment techniques to attain an accurate parameter estimation. We study a real experimental device devoted to evaluate the effect of temperature and light on microalgae growth. On the basis of a mathematical model of the experimental system, the optimal experiment design problem was formulated and solved with both static (constant light and temperature) and dynamic (time varying light and temperature) approaches. Simulation results indicated that the optimal experiment design allows for a more accurate parameter estimation than that provided by the existing experimental protocol. For its efficacy in terms of the maximum likelihood properties and its practical aspects of implementation, the dynamic approach is recommended over the static approach.

Keywords: biofuel; biological processes; modelling; parameter identification; optimal experiment design

Email addresses: rafaun@yahoo.com (Rafael Muñoz-Tamayo), pierre.martinon@inria.fr (Pierre Martinon), gael.bougaran@ifremer.fr (Gaël Bougaran), francis.mairet@inria.fr (Francis Mairet), olivier.bernard@inria.fr (Olivier Bernard)

1. Introduction

Microalgae have received a specific attention in the framework of renewable energy generation [1]. However, optimizing productivity in large scale systems is a difficult task since microalgae growth is driven by multiple factors including light intensity, temperature and pH [2]. Mathematical modelling is thus required for quantifying the effect of environmental factors on microalgae dynamics.

In order to obtain reliable models that can be used in prediction and optimization of large scale systems, the model calibration stage requires carefully designed experiments with high informative content. Providing accurate parameters is indeed crucial since model-based optimality might be sensitive to parameters values as shown in [3]. Moreover, assessing the effect of operational factors via sensitivity analysis can provide useful information for improving configuration design of photobioreactors [4].

Looking for high informative experiments is the objective of optimal experiment design (OED) for parameter estimation. Extensive work has been done for tackling the OED problem for dynamical systems (see, *e.g.*, [5, 6, 7, 8]). The OED problem can be formulated as an optimal control problem. For low dimension models, analytical solutions may be obtained by the application of Pontryagin's Maximum Principle, which provides necessary conditions to be satisfied by the optimal inputs (see,*e.g.* [9]). When model complexity increases, analytical solutions are arduous to obtain and thus the solution of the OED problem relies on numerical optimization techniques (see,*e.g.* [10, 11]).

When dealing with biological systems, OED approaches are based either on static or dynamic experiments (see, *e.g.*, [12, 13]). In this work, we analyze these two strategies and capitalize the available tools for OED to provide guidelines for the design of optimal experiments that allow an efficient assessment of the effect of temperature and light on microalgae growth. The model under investigation represents a real experimental device used to assess optimal growth conditions under batch mode. This device is operated at IFREMER Nantes, France.

The paper is organized as follows. Section 2 presents the system under study and its

mathematical description, which corresponds to a simplified model of microalgae growth. The OED framework based on this model is detailed in Section 3. In Section 4 we show the results of solving the OED problem. Two strategies are analyzed, namely static and dynamic approaches. Furthermore, we discuss about the relevance of OED for model-driven decisions on raceway performance. For that, we make use of a local sensitivity analysis of a more complex model describing microalgae growth in an outdoor pond. In Appendix A, we discuss about the structural and practical identifiability of the model. The main conclusions of the study are summarized in Section 5.

2. Modelling

We focus our study on the effect of temperature and light on the growth of microalgae. More precisely, we aim at designing efficient experimental protocols for a real experimental system that allow an accurate estimation of the model parameters. The experimental apparatus, named the TIP (Fig. 1), consists of 18 batch photobioreactors located inside an incubator (see [14] for more details). In each photobioreactor, it is possible to regulate the temperature, pH and light intensity.

Following the models developed for microalgae growth [15, 16], we study here a simplified model of microalgae growth under the hypotheses that the experiment is carried out at low cellular concentrations and under conditions of non-limiting nutrients. The first hypothesis implies that light is homogeneous along the depth of the photobioreactor. The second hypothesis implies that the cells grow in exponential phase. The resulting mass balance equation on the TIP system reads

$$\dot{x} = f(x, \boldsymbol{\theta}, I, T, t) = \overline{\mu}(\boldsymbol{\theta}, I, T)x, \;\; x(0) = x_0, \tag{1}$$

with x the biomass concentration, I the light intensity and T the temperature in the reactor, $\boldsymbol{\theta}$ the parameter vector and $\overline{\mu}(\cdot)$ the specific growth rate $\overline{\mu}(\cdot)$ defined by

$$\overline{\mu}(\boldsymbol{\theta}, I, T) = \mu_{\max}\phi_I\phi_T. \tag{2}$$

with $\mu_{\max}$ the maximal specific growth rate. The factors ϕ_I, ϕ_T, detailed below, represent the effects of light and temperature on microalgae growth.

Temperature has a homogeneous effect on uptake and growth rates [17]. The effect of temperature is described by the cardinal model developed for bacteria [18] and validated for microalgae [16].

$$\phi_T = \begin{cases} 0, & T < T_{\min} \\ \frac{(T-T_{\max})(T-T_{\min})^2}{(T_{\text{opt}}-T_{\min})[(T_{\text{opt}}-T_{\min})(T-T_{\text{opt}})-(T_{\text{opt}}-T_{\max})(T_{\text{opt}}+T_{\min}-2T)]}, & T \in [T_{\min}, T_{\max}] \\ 0, & T > T_{\max}. \end{cases} \tag{3}$$

The effect of light (ϕ_I) on microalgae growth is often represented by a Haldane type kinetics that accounts for photoinhibition [19]. The following parameterization of the standard Haldane equation is used [16]

$$\phi_I = \frac{I}{I + \frac{\mu_{\max}}{\alpha}\left(\frac{I}{I_{\text{opt}}} - 1\right)^2}, \tag{4}$$

where α is the initial slope of the growth response curve w.r.t. light.

In terms of practical identifiability properties, Eq. (4) excels the standard Haldane kinetics. For a brief discussion, the reader is referred to Appendix A.

The above equations implies that microalgae exhibit a maximal growth rate at optimal conditions of light (I_{opt}) and temperature (T_{opt}).

The model is then determined by the parameter vector $\boldsymbol{\theta}$

$$\boldsymbol{\theta} = [\mu_{\max},\ \alpha,\ I_{\text{opt}},\ T_{\min},\ T_{\max},\ T_{\text{opt}}].$$

In the next Section, we tackle the OED problem locally, that is the design of optimal experiments is carried out on the basis of nominal values $\hat{\boldsymbol{\theta}}$. Table 1 shows the nominal values of the model parameters used in this study. They correspond to the microalgae *Isochrysis* aff. *galbana*, currently named as *Tisochrysis lutea* [20]. Parameter values were mainly obtained from [3] and [21]. The temperature parameters are those of *Nannochloropsis oceanica* [16] whose maximal and optimal temperatures are close to those of *Tisochrysis lutea* [22].

3. OED problem for parameter estimation

The problem of OED for parameter estimation consists in designing an experimental protocol that provides data with high informative content to allow an accurate identification of the model parameters, that is to provide estimates with small confidence intervals. Classical approaches of OED for parameter estimation rely on the optimization of a scalar function of the Fisher information matrix (FIM), since this matrix is the core for the calculation of the confidence intervals of the parameter estimates (see, *e.g.*, [6], [8]). Recent approaches such as the Sigma Point Method have been proposed to estimate parameter uncertainty without the explicit calculation of the FIM [23]. Here, we will focus on the classical approach.

Let us recall some basic principles of parameter identification. We consider here a local design approach. Our aim is to design optimal experiments on the basis of the nominal parameter vector $\hat{\boldsymbol{\theta}}$. We first assume that the ith measurement (observation) y_i of our experiment is modelled as:

$$y_i = y_{\mathrm{m}_i}(\boldsymbol{\theta}^*) + \varepsilon, \tag{5}$$

where $y_{\mathrm{m}_i}(\boldsymbol{\theta}^*)$ is the deterministic output of the model and $\boldsymbol{\theta}^*$ the true value of the parameter vector. The measurement error ε is here assumed to follow a normal distribution $\varepsilon \sim \mathbf{N}(0, \sigma^2)$. Note that (5) implies that a deterministic model is available and represents adequately the system. Moreover, the model structure must be structurally identifiable. In Appendix A, structural identifiability of the model is checked.

The maximum likelihood (ML) estimate $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ minimizes the cost function

$$J(\boldsymbol{\theta}) = \frac{1}{\sigma_{\mathrm{s}}^2} \sum_{i=1}^{n} \left(y_i - y_{\mathrm{m}_i}(\boldsymbol{\theta})\right)^2, \tag{6}$$

with n the number of data points.

The covariance matrix $\widehat{\mathbf{P}}$ of $\widehat{\boldsymbol{\theta}}$ can be approximated to

$$\widehat{\mathbf{P}} = \mathbf{F}^{-1}(\widehat{\boldsymbol{\theta}}), \tag{7}$$

where $\mathbf{F}$ is the Fisher information matrix. An estimate of the standard deviation of $\widehat{\theta}_j$ is given by

$$\eta_j = \sqrt{\widehat{\mathbf{P}}_{jj}}. \tag{8}$$

We will be then interested in designing an experiment that render η_j small.

In our case study, we aim at determining optimal profiles (or levels) of temperature (T) and light intensity (I) for attaining an accurate estimation of parameters. Optimal experiments are built w.r.t. the D-optimality criterion, which maximizes the determinant of the FIM. Maximizing the determinant implies minimizing the volume of the confidence ellipsoids for the parameters [6].

By means of simulations, we tested also other optimality criteria, namely E-optimality (maximization of the smallest eigenvalue of the FIM) and modified E-optimality ((minimization of the condition number of the FIM)). D-optimality provided the best results in terms of the volume of the confidence ellipsoids. Therefore, we chose D-optimality as criterion of optimal design. Interestingly, the modified E-optimality criterion resulted in large confidence intervals. Indeed, it has been noted that since the modified E-optimality criterion is a criterion of shape of the ellipsoids, it is possible to obtain circular confidence regions with large volumes [24].

It should be noted that the performance of the obtained optimal experiment strongly depend on the nominal values of the estimates of the parameter vector. Ideally, $\widehat{\boldsymbol{\theta}}$ should be as close as possible to $\boldsymbol{\theta}^*$. In our case study, the nominal values of the parameters used are expected to be close to the true values, since the selection of *priors* was based on published experimental studies.

The OED problem is tackled by means of two strategies, namely dynamic and static approaches, which are detailed in the following.

3.1. Dynamic approach

The OED by the dynamic approach is directly applied on the dynamic (primary) model (1). Here, the temperature and light intensity can be set to vary in time.

For the dynamic approach, the FIM reads as follows

$$\mathbf{F}_{\mathrm{d}}(\widehat{\boldsymbol{\theta}}) = \frac{2}{\sigma_{\mathrm{d}}^2} \sum_{k=1}^{n_{\mathrm{e}}} \sum_{i=1}^{n_{\mathrm{t}}} \left[\frac{\partial y_{\mathrm{m}_{k,i}}}{\partial \boldsymbol{\theta}} \right]_{\widehat{\boldsymbol{\theta}}}^{\mathrm{T}} \left[\frac{\partial y_{\mathrm{m}_{k,i}}}{\partial \boldsymbol{\theta}} \right]_{\widehat{\boldsymbol{\theta}}} = \frac{2}{\sigma_{\mathrm{d}}^2} \widehat{\mathbf{M}}_{\mathrm{d}}, \tag{9}$$

with n_{t} the number of sampling times. Here, $y_{\mathrm{m}_{k,i}}$ is the biomass concentration predicted by the model (1) for the kth experiment at the ith time and σ_{d}^2 is the noise variance of the

measurement of biomass concentration. $\widehat{\mathbf{M}}_{\mathrm{d}}$ is the matrix resulting from the summation term. This formulation is made for facilitating further discussion. The terms in brackets in (9) contains the local sensitivities of the model output w.r.t. the parameter vector $\boldsymbol{\theta}$. The sensitivity functions were computed automatically with the Matlab Toolbox IDEAS [25]. The toolbox is devoted to estimate parameters of ODE models. It uses symbolic differentiation to calculate the sensitivity functions for the evaluation of the FIM.

An approximate noise variance $\sigma_{\mathrm{d}}^2 = 9.31$ was calculated from the data reported in [26] and the mathematical model developed in [27].

The OED problem is defined as

$$\min_{\boldsymbol{\varphi}_{\mathrm{d}}} \mathrm{Det}\left(\mathbf{F}\right), \tag{10}$$

with $\boldsymbol{\varphi}_{\mathrm{d}}$ the design vector

$$\boldsymbol{\varphi}_{\mathrm{d}} = [T_1(t), I_1(t), \ldots, T(t)_{n_{\mathrm{e}}}, I(t)_{n_{\mathrm{e}}}],$$

such that

$$\begin{aligned} T_{\mathrm{L}} = 12 \leq T_k(t) \leq T_{\mathrm{U}} = 33.2^{\circ}\mathrm{C} \\ I_{\mathrm{L}} = 20 \leq I_k(t) \leq I_{\mathrm{U}} = 1200\ \mu\mathrm{E\ m^{-2}s^{-1}} \\ \dot{T}_{\mathrm{L}} = -5 \leq \dot{T}_k(t) \leq \dot{T}_{\mathrm{U}} = 15^{\circ}\mathrm{C}, \end{aligned} \tag{11}$$

with n_{e} the number of distinct experiments. We set $n_{\mathrm{e}} = 9$ with duplicate experiments. The boundaries in (11) correspond to the physical boundaries of the TIP system. Note that the rate of temperature change ($\dot{T}$) is imposed. This constraint is bounded by the thermal dynamics of the equipment but also it must account for the potential thermal stress induced to the microalgal cells.

No boundaries were imposed to the rate of change of light, since it can be changed instantaneously. We assumed that microalgae respond instantaneously to light changes. However, it is known that microalgae can adapt its photosynthetic system to changes of light [17]. Here, we consider time scales larger than the photosynthesis response time (in the range of minutes for photoinhibition). In this case study, we neglect however photoacclimation (adaptation of the pigment content to light intensity, at the scale of

weeks). Further experiments will be needed to assess the dynamics of such an adaptation.

Note that $\boldsymbol{\varphi}_{\mathrm{d}}$ is of infinite dimension. However, $\boldsymbol{\varphi}_{\mathrm{d}}$ will be further transformed into a finite dimension vector to solve the optimization problem numerically.

Before attempting to solve the full OED problem, we first partitioned the original OED problem into simpler subproblems in which we study the effect of either temperature or light. This strategy was for instance used in [28] to estimate the cardinal temperatures for *E. coli*.

Each subproblem is dedicated to improve the accuracy of the estimation of a couple of parameters, while the other parameters are assumed to be known. In this case, $\mathbf{F}_{\mathrm{d}}$ is a square matrix of dimension 2×2 for each subproblem (for the full OED problem, the FIM is of dimension 6×6). The initial concentration of biomass was set to $x_0 = 10$ mg/L and. The duration of the experiment to $t_f = 4$ d with ten equidistant sampling times.

When studying the temperature parameters, the light was set to $I = 547$ μE m$^{-2}s^{-1}$, and when studying the light, the temperature was set to $T = 26.7$ °C. These constant values correspond to the optimal values for growth obtained from the model parameters (Table 1). This choice is supported by the fact that the FIM of each subproblem only involves either parameters related to the effect of light or to the effect of temperature, therefore the other experimental input only affects relatively the calculation of the sensitivity functions. By setting the experimental inputs to their optimal values, we favor growth.

A total number of nine subproblems was obtained. Table 2 shows the combination of parameters and the experiment input (T or I) for each subproblem. In practice, the nine solutions will be implemented in duplicates in the TIP.

The resulting subproblems were solved numerically via two discretization methods, namely sequential and simultaneous. The discretization allows to convert the original infinite dimensional optimization problem into a finite dimension problem. In the sequential approach (also called control vector parametrization (CVP)), the control variables are approximated by a set of basis functions that depend on a finite number of real parameters. In the simultaneous approach, all state and control variables are discretized w.r.t. time. Hence, this method is also known as total discretization. In this case, the dimension of

the optimization problem depends on the number of discretization steps [29, 30].

The simultaneous method was implemented with the open source toolbox BOCOP [31](`http://bocop.org`), based on the IPOPT solver [32]. The simultaneous method used a Midpoint discretization with 1000 steps, with a 10^{-14} tolerance for solving the discretized problem. All state and control variables were initialized with constant values.

Numerical solutions of the CVP approach were obtained with the SSMGO toolbox (`http://www.iim.csic.es/ gingproc/ssmGO.html`), with the parameterization depicted in Fig. 2. SSmGo performs global optimization by using a scatter search method [33, 34].

The experiment inputs are thus defined by four parameters, namely u_1, u_2, t_1, t_2. The dimension of the optimization problem is therefore 9×4 with the decision vector

$$\boldsymbol{\varphi}_{\mathrm{d}} = [u_1(1), u_2(1), t_1(1), t_2(1) \ldots, u_1(n_{\mathrm{e}}), u_2(n_{\mathrm{e}}), t_1(n_{\mathrm{e}}), t_2(n_{\mathrm{e}})]\,. \tag{12}$$

3.2. Static approach

The OED by the static approach is based on the secondary model of growth (here represented in (2)). In this approach, one experiment is characterized by a constant environment (T,I in our case). The dynamic data of the biomass evolution for a given experiment is first used to calculate the maximal growth. Once different growth rates are calculated at different conditions of temperature and light intensity, the parameter estimation procedure is applied on the growth model (2).

Since the TIP system allows to run 18 experiments simultaneously, a parallel design procedure is here used. Hence, the following OED strategies aim at finding the nine best experiment conditions to account for duplicate experiments.

For the static approach, the FIM is computed as

$$\mathbf{F}_{\mathrm{s}}(\widehat{\boldsymbol{\theta}}) = \frac{2}{\sigma_{\mathrm{s}}^2} \sum_{k=1}^{n_{\mathrm{e}}} \left[\frac{\partial y_{\mathrm{m}_k}}{\partial \boldsymbol{\theta}}\right]_{\widehat{\boldsymbol{\theta}}}^{\mathrm{T}} \left[\frac{\partial y_{\mathrm{m}_k}}{\partial \boldsymbol{\theta}}\right]_{\widehat{\boldsymbol{\theta}}} = \frac{2}{\sigma_{\mathrm{s}}^2} \widehat{\mathbf{M}}_{\mathrm{s}}, \tag{13}$$

where y_{m_k} is the maximal growth predicted by the model (2) for the kth experiment, n_{e} is the number of distinct experiments ($n_{\mathrm{e}} = 9$) and σ_{s}^2 is the noise variance associated to measurement of the maximal growth. To provide an approximate value of the noise

variance, the dynamic model was simulated for nine experiments. Each of them characterized by a level of temperature and light intensity. Normal distributed data of biomass concentration was further generated by taking the value of noise variance of biomass σ_{d}^2. The generated noisy data was used to calculate the variance of specific growth. An approximate value of $\sigma_{\mathrm{s}}^2 = 3.8 \cdot 10^{-3}$ was obtained.

The OED problem is then defined as

$$\min_{\boldsymbol{\varphi}_{\mathrm{s}}} \mathrm{Det}\left(\mathbf{F}\right), \tag{14}$$

with $\boldsymbol{\varphi}_{\mathrm{s}}$ the design vector

$$\begin{aligned} \boldsymbol{\varphi}_{\mathrm{s}} &= \left[T_1, I_1, \ldots, T_{n_{\mathrm{e}}}, I_{n_{\mathrm{e}}}\right], \\ &\text{such that} \\ T_{\mathrm{L}} = 12 \leq T_k &\leq T_{\mathrm{U}} = 33.2^{\circ}\mathrm{C} \\ I_{\mathrm{L}} = 20 \leq I_k &\leq I_{\mathrm{U}} = 1200\ \mu\mathrm{E\ m^{-2}s^{-1}}. \end{aligned} \tag{15}$$

The design vector $\boldsymbol{\varphi}_{\mathrm{s}} \in \mathbb{R}^{n_{\mathrm{e}}}$.

To solve the OED problem of the static approach, the Matlab optimization toolbox SSmGo was used.

4. Results and Discussion

Before presenting the resulting optimal experiments for both static and dynamic approaches, we should keep in mind that in our case study the D-optimal experiments do not depend on the value of the noise variance σ^2, given that we assumed that the measurement errors are homoscedastic. Indeed, the optimal experiment inputs depend only on the matrix $\widehat{\mathbf{M}}$, defined previously in (9,13). On the other hand, the confidence intervals of the estimates do depend on the actual value of σ since the estimate of the standard deviation of the parameter θ_j is given by

$$\eta_j = \frac{\sigma}{\sqrt{2}} \sqrt{(\widehat{\mathbf{M}}_{jj})^{-1}}. \tag{16}$$

4.1. OED by the static approach

The nine D-optimal experiments are given in Table 3. These experiments are defined by six levels of light intensity and five levels of temperature (if the decimal digits are omitted). Note that the nine experiments include the repetition of two experimental conditions (experiments 1,2 and experiments 5,6), which results in seven distinct experimental conditions. This result is not surprising since D-optimal often calls for the repetition of a small number of experimental conditions [6]. Simulated data resulted from the D-optimal experiments are illustrated in Fig 5A.

The performance of the D-optimal experiments was compared by means of simulation with a equidistant full 3^2 factorial design including duplicates and with the central composite design currently used in the TIP device [14]. This composite design involved 17 experiments with five levels for the environmental variables temperature, pH and light intensity. Since in our study the effect of the pH is not considered in the OED problem, we only took into account the levels for temperature and light of the 17 experiments. The maximal level of temperature used in [14] was 33.7°C. We set the maximal temperature of culture to 33.2°C, which is lower than the nominal value of the upper temperature for algae growth ($T_{\max}$).

Table 4 illustrates the advantage of the D-optimal experiments over the factorial designs. Firstly, we notice that with the equidistant full factorial design the determinant of the FIM is zero, implying that the FIM is singular. Indeed, the inverse of the condition number of the FIM (defined as ratio of the largest eigenvalue to the smallest one) is smaller than the precision of floating point format ($2{\cdot}10^{-16}$). In this case, confidence intervals for the parameter estimates can not be computed on the basis of the density of the estimator. To identify alternatives for guaranteeing a non-singular FIM for a full factorial design, a series of computations was performed. From the computations, it is concluded that a miminum number of four levels need to be considered in a full factorial design to provide a well-conditioned FIM. Other option to avoid an ill-conditioned FIM is to reduce the dimension of the matrix by splitting the full problem into subproblems (as we did for the dynamic OED). Our computations indicated that for combinations of five parameters $\binom{6}{5}$, five out of six possible combinations of parameters led to a well-conditioned FIM.

The combination that resulted in a singular FIM was [$\mu_{\max}$, α, $T_{\min}$, $T_{\max}$, T_{opt}]. For combinations of four parameters (FIM has dimension 4×4), the FIM was well-conditioned for all the fifteen combinations.

It should be noted that in a simulation study performed in [35], full factorial design was applied for a cardinal model describing the effects of temperature, pH and water activity on the microbial growth rate, and the estimated values were close to the nominal values used in the simulation. However, we should not be content only with this result, since the actual values need to be supported by their corresponding confidence intervals in order to identify practical identifiability problems and to provide a quantitative measurement of the accuracy of the estimation.

We note that the composite factorial design does provide a well-conditioned FIM. However the determinant of the FIM for this design is much lower than that obtained with the D-optimal design, and this is actually reflected on the accuracy of the estimates. The second raw of Table 4 shows the ratio of the standard deviations of the parameters obtained with the D-optimal design to the standard deviations obtained with the composite design. D-optimal design provides lower standard deviations, 36% better in average. This result establishes the benefit of designing optimal experiments with OED techniques for obtaining accurate parameter estimates.

4.2. OED by the dynamic approach

As it was mentioned in Section 3, the dynamic OED problem (10-11) was solved via the simultaneous and CVP approaches. While the CVP method reduces substantially the dimension of the original optimization problem, the simultaneous approach allows to find solutions without restricting the shape of the controls. These solutions potentially give better objective values, but may not be fit for practical use, if the controls have a complicated shape. Comparing the two methods also give a hint at what we lose by restricting the control shape to simple functions.

The CVP and the simultaneous approaches were compared for the case when the full problem was partitioned into nine subproblems devoted to improve the accuracy of the estimation of a couple of parameters. Overall, the CVP and the simultaneous methods

find very similar solutions with the exception of the experiment 7 (see Fig. 3). The controls found with the simultaneous approach are often quite close in shape to piecewise linear functions. This confirms that our choice of shape for the controls in the CVP was a sensible one.

Table 5 compares the optimality cost functions provided by the simultaneous (J_{Sim}) and the CVP (J_{cvp}) methods. For all the nine experiments, the simultaneous approach converges to better solutions than the CVP ones. However, the CVP approach provides optimality cost functions very close to those obtained with the simultaneous approach. In average, the cost functions obtained by the CVP approach are 95% of those obtained with the simultaneous approach.

From the study of the subproblems, we can conclude that the CVP approach with a simple piecewise parametrization seems well suited to design highly informative experiments. We now apply the CVP approach to the full OED problem, with the FIM of dimension 6×6. The optimal experiment inputs obtained are displayed in Fig. 4. The simulation of the nine D-optimal experiments is displayed in Fig 5B. Note that in the experiment 9, the biomass concentration exhibits, for a certain time interval, a behavior close to the steady state. This is due to the fact that the temperature reaches a very close value to $T_{\max}$ and thus the growth rate becomes close zero. When performing the experiments, caution should be made for the selection of the maximal operational temperature. Indeed, an erroneous *a priori* on $T_{\max}$ with a higher value than the real maximal temperature would lead to cell inactivation [13]. For microalgae cultures, $T_{\max}$ must be well characterized to avoid operations that can be detrimental for attaining maximal productivities [38]. In our case study, we were conservative in the selection of the *prior* of $T_{\max}$. By setting the *prior* lower than the maximal value reported in [14], we assured that the temperature will allow growth in all the experiments.

Additionally, we wanted to assess the accuracy of the estimates when applying the optimal solutions obtained from the nine small subproblems to the full OED. Table 6 shows the ratio of the standard deviations of the estimates obtained from the full OED solutions to those obtained from the solutions of the OED subproblems. We observe that the standard deviations obtained when solving the full OED problem are usually smaller than

those for the partitioned subproblems, 70% in average. For $T_{\max}$ the estimated standard deviations of the two approaches are very close. For the light associated parameters, in particular, the accuracy of the estimation provided by the solution of the full OED problem is substantially better. A higher determinant of the FIM (two orders of magnitude) is obtained with the full OED solution while the condition numbers for the partitioned and the full OED are of the same magnitude. On the other hand, we also see that partitioning the full OED problem into small subproblems gives satisfactory results. This strategy of simplification could be easier to implement when dealing with the full model, even if a better accuracy is achieved by considering the full FIM.

4.3. Static vs Dynamic OED

A complete comparison between the static and the dynamic approaches for OED would require the knowledge on the noise variance for the measurements of the maximal growth rate (σ_{s}^2) and the biomass concentration (σ_{s}^2). However, even if this information is unknown *a priori* we can still draw a comparative analysis of the performance of these methods, assuming that the data is generated by (5).

The unbiased estimator of the noise variance reads as

$$\sigma^2 = \frac{1}{n - n_{\mathrm{p}}} \sum_{i=1}^{n} [y_i - y_{\mathrm{m}_i}(\boldsymbol{\theta}^*)]^2, \tag{17}$$

with n the total number of data measurements and n_{p} the number of parameters . Since the ML estimator is efficient asymptotically (as $n \to \infty$), we can infer that the dynamic approach provides a more efficient estimator than the static approach. Indeed, for our case study, while the number of data points in the static approach is only 3 times the number of parameters, when applying the dynamic approach we get a number of experimental points that is 30 times the number of parameters.

To allow for a quantitative comparison, we used the approximated noise variances previously estimated $\sigma_{\mathrm{d}}^2 = 9.31$ and $\sigma_{\mathrm{s}}^2 = 3.8 \cdot 10^{-3}$ to generate random simulated data for tackling the parameter estimation problem for both methods. Table 7 shows the estimated values and their confidence intervals for both approaches. The standard deviations of the parameters obtained with the dynamic approach are in average 42% lower than those given by the static approach. For the parametes $\mu_{\max}, I_{\mathrm{opt}}$ the dynamic approach excels

substantially the static approach by providing standard deviations 13% lower. Finally, it is worth noting that for the static approach to equal in average the dynamic approach, it is required to reduce significantly the value of σ_s, which is only possible for $n >> n_p$.

The correlation matrix of the estimates for the dynamic approach was

$$\begin{array}{cccccc}
\mu_{max} & 1.0 & & & & \\
\alpha & -0.47 & 1.0 & & & \\
I_{opt} & -0.08 & -0.36 & 1.0 & & \\
T_{min} & 0.34 & -0.04 & -0.07 & 1.0 & \\
T_{max} & -0.19 & 0.10 & -0.04 & 0.33 & 1.0 \\
T_{opt} & -0.14 & 0.01 & 0.05 & -0.35 & -0.34 & 1.0
\end{array}$$

For the static approach, The correlation matrix of the estimates was

$$\begin{array}{cccccc}
\mu_{max} & 1.0 & & & & \\
\alpha & -0.47 & 1.0 & & & \\
I_{opt} & -0.32 & 0.32 & 1.0 & & \\
T_{min} & 0.49 & -0.19 & -0.15 & 1.0 & \\
T_{max} & -0.07 & 0.03 & 0.03 & 0.18 & 1.0 \\
T_{opt} & -0.21 & 0.04 & -0.02 & -0.39 & -0.33 & 1.0
\end{array}$$

The condition numbers of both approaches are comparable (see Table 4 and Table 6). The correlation matrices for both approaches indicated a low correlation between the parameters despite the high condition numbers . This is indeed thanks to the practical identifiability properties of the cardinal model as discussed in Appendix A.

For the dynamic approach the mean squared error (MSE) of the estimated parameters w.r.t the nominal parameters was 1.42 while for the static approach MSE = $2.30{\cdot}10^3$, indicating the dynamic approach provides closest estimates to the nominal values in comparison to the static approach. Only for T_{max}, both approaches perform equally.

Practical aspects as the labor of performing a two-step identification [12] place the static approach in disadvantage compared to the dynamic approach. These reasons lead us to favor the dynamic approach. Another benefit is that the sampling times could be further optimized within the experimental protocol, giving additional degrees of freedom.

From the mechanistic point of view, by stimulating the system with time-varying inputs, the dynamic approach allows a better characterization of the system behavior. On the opposite, the static approach can hide the relevance of certain important phenonema. This factor is critical to our case study where microalgae are meant to grow in a dynamic environment that is periodically forced by daily variations of light and temperature. However, we should keep in mind that to take advantage fully of the dynamic approach, a step forward in the mathematical description of the process needs to be done for accounting important phenomena such as acclimation to light and temperature [17, 36] and cell inactivation due to high temperatures. For the sake of generality, a further study should be done to include the impact of the initial conditions and the physiological state of the cells on the determination of optimal experiment inputs. We also recommend to perform a preliminary experiment for which the cells get acclimated to their light and temperature growth conditions. This experiment will allow a dynamic characterization of the adaptation phenomena.

4.4. Relevance of accurate estimation on model-driven optimization

One of the ultimate goals of developing microalgae growth models is to provide a platform for model predictions and for the design of optimal control strategies for systems operated at large scale. Following this aim, we wanted to assess the relevance of providing accurate parameter estimates on the quality of the predictions for a more complex model representing the continuous cultivation of microalgae on an outdoor pond. For that, we used the raceway model described in [3]. The model takes the configuration of a pilot-scale raceway (Algotron) located at INRA LBE, Narbonne (France). The model is described by

$$\dot{s} = D(s_{in} - s) - \rho(\cdot)x, \tag{18}$$

$$\dot{q_n} = \rho(\cdot) - (\mu(\cdot) - R(\cdot))q_n, \tag{19}$$

$$\dot{x} = (\mu(\cdot) - D - R(\cdot))x, \tag{20}$$

where s (mg N/L) is the extracellular nitrogen concentration and x (mg C/L) is the concentration of carbon biomass. The term q_n (g N/g C) denotes the intracellular nitrogen

quota, that is the concentration of nitrogen per biomass unit. D is the dilution rate, $\mu(\cdot)$ is the specific growth rate, ρ is the nitrogen uptake rate and $R(\cdot)$ the respiration rate. For more details, the reader is referred to [3].

Firstly, we evaluated the sensitivity of the biomass concentration with respect to the model parameters along a year of operation. Meteorological data was used for the location of Narbonne to calculate the temperature and light intensities for each month. The normalized sensitivity matrix $\mathbf{S}_y$ was computed for each month. The element (k, j) of $\mathbf{S}_y$ is calculated as [37]

$$\mathbf{S}_y(k,j) = \sum_{i=1}^{n_\mathrm{t}} \left| \overline{s}_j^k(t_i, \widehat{\boldsymbol{\theta}}) \right|, \tag{21}$$

where $\overline{s}_j^k$ is the normalized sensitivity of the model output y_{m_k} w.r.t. θ_j,

$$\overline{s}_j^k(t_i, \widehat{\boldsymbol{\theta}}) = \frac{\widehat{\theta}_j}{y_{\mathrm{m}_k}(t_i, \widehat{\boldsymbol{\theta}})} \left[\frac{\partial y_{\mathrm{m}_k}}{\partial \theta_j} \right]_{(t_i, \widehat{\boldsymbol{\theta}})}. \tag{22}$$

Figure 6 shows a graphical representation of the sensitivity matrices for four months. January is the coldest month in Narbonne, while August is the warmest. October is an intermediate month. The sensitivities of June are also presented for illustration. It is interesting to observe that the influence of the parameters on the model response is modulated by the environmental conditions. Indeed, we can see a specific pattern of parameter influence for each month. In terms of the tuning importance, that is the importance of parameter changes around their nominal value for the model output [37], we observe that, overall, the most dominant parameter is T_{opt}. In August the most dominant parameter is $\mu_{\max}$. This month exhibits the most homogeneous distribution of the influence of parameters in comparison with months like January where the influence of two parameters $(T_{\max}, T_{\mathrm{opt}})$ exceeds substantially the influence of the rest of parameters. In cold months (*e.g.* January-March), the influence of $T_{\max}$ is higher than the influence of $\mu_{\max}$. This pattern is switched in warm months (*e.g.* June-August). In the figure, $T_{\min}$ appears as the less influential parameter. This effect may be inverted in cold regions. Indeed, the minimum average temperature in Narbonne used in our simulation is 4.76 °C, which is very high compared to the nominal value of $T_{\min}$.

Figure 7 shows the dramatic effect of an uncertainty of 5% on the nominal value of T_{opt} on the quality of model predictions for the month of January. Both overestimation ($1.05T_{\text{opt}}$) and underestimation ($0.95T_{\text{opt}}$) of the optimal temperature results in important discrepancies between the response of the model with the nomical value of T_{opt} and those obtained with a small perturbation of 5% on the nominal value. Hence the importance of providing accurate estimates since small changes on the parameter values can induce large changes on the biomass dynamics. Model-driven decisions are thus strongly dependent on the accuracy of the parameter estimates.

The previous result strengthens the relevance of the temperature effect for outdoor cultivation as discussed in [38]. It should be noted that with the meteorological data used here, the temperature of the culture (T) never exceeded T_{max}, so the effect of temperature ϕ_{T} was always higher than zero. We recalled that an overestimation on T_{max} will have a strong impact on model predictions and system operation. In particular, when the temperature exceeds T_{max} phenonema as cell inactivation and mortality take place. These phenomena, detrimental for attaining maximal productivities, need to be characterized by an approach combining both experiments and modelling in order to provide guidelines to mitigate negative effects.

5. Conclusions

We solved the OED problem for a simplified model of microalgae growth. We have determined optimal experiment conditions to provide an accurate estimation of the parameters that drive microalgae growth by modulating the effects of light and temperature. Both static and dynamic approaches were evaluated to find D-optimal experiments. From our results, we recommend the use of the dynamic approach in virtue of the efficacy in terms of the maximum likelihood properties of the estimator. The protocol of experiment inputs determined in this study will be further implemented in the TIP system used at Ifremer Institute.

For the dynamic case, we showed that a parameterization of the control input by piecewise linear functions (CVP approach) provides efficient results as compared as the simultaneous approach. Moreover, the strategy of partitioning the full OED problem

into subproblems dedicated to improve the accuracy of the estimation of a couple of parameters was shown to be satisfactory. The CVP method and the partitioning of the full OED into subproblems are suitable approaches for solving the OED problem in microalgae growth models by reducing, additionally, the problem complexity. Finally, with the use of sensitivity analysis of a more complex model describing the cultivation of microalgae in a raceway, we showed the relevance of providing accurate parameters for enabling reliable model-driven decisions.

6. Acknowledgment

This work benefited from the support of the Facteur 4 research project founded by the French National Research Agency (ANR).

References

[1] P. J. B. Williams, L. M. L. Laurens, Microalgae as biodiesel & biomass feedstocks: Review & analysis of the biochemistry, energetics & economics, Energy and Environ. Sci. 3 (2010) 554–590.

[2] C. Posten, Design principles of photo-bioreactors for cultivation of microalgae, Eng. Life Sci. 9 (2009) 165–177.

[3] R. Muñoz-Tamayo, F. Mairet, O. Bernard, Optimizing microalgal production in raceway systems., Biotechnol Prog 29 (2013) 543–552.

[4] D. A. Pereira, V. O. Rodrigues, S. V. Gómez, E. A. Sales, O. Jorquera, Parametric sensitivity analysis for temperature control in outdoor photobioreactors, Bioresource technology 144 (2013) 548–553.

[5] G. Goodwin, R. Payne, Dynamic System Identification: Experiment Design and Data Analysis, Academic Press, New York, 1977.

[6] E. Walter, L. Pronzato, Identification of Parametric Models from Experimental Data, Springer, London, 1997.

[7] K. Keesman, System Identification: an Introduction, Springer-Verlag, London„ 2011.

[8] G. Franceschini, S. Macchietto, Model-based design of experiments for parameter precision: state of the art, Chemical Engineering Science 63 (19) (2008) 4846–4872.

[9] J. Stigter, K. Keesman, Optimal parametric sensitivity control of a fed-batch reactor, Automatica 40 (2004) 1459 – 1464.

[10] I. Bauer, H. G. Bock, S. Körkel, J. P. Schlöder, Numerical methods for optimum experimental design in DAE systems, Journal of Computational and Applied Mathematics 120 (2000) 1–25.

[11] E. Balsa-Canto, A. A. Alonso, J. Banga, Computational procedures for optimal experimental design in biological systems, IET Syst. Biol. 2 (2008) 163–172.

[12] K. Bernaerts, K. J. Versyck, J. F. Van Impe, On the design of optimal dynamic experiments for parameter estimation of a Ratkowsky-type growth kinetics at suboptimal temperatures, International journal of food microbiology 54 (1) (2000) 27–38.

[13] K. Bernaerts, K. P. M. Gysemans, T. Nhan Minh, J. F. Van Impe, Optimal experiment design for cardinal values estimation: guidelines for data collection., Int J Food Microbiol 100 (2005) 153–165.

[14] J. Marchetti, G. Bougaran, L. L. Dean, C. Mégrier, E. Lukomska, R. Kaas, E. Olivo, R. Baron, R. Robert, J. P. Cadoret, Optimizing conditions for the continuous culture of *Isochrysis* affinis *galbana* relevant to commercial hatcheries, Aquaculture 326-329 (2012) 106–115.

[15] O. Bernard, Hurdles and challenges for modelling and control of microalgae for CO2 mitigation and biofuel production, Journal of Process Control 21 (2011) 1378–1389.

[16] O. Bernard, B. Remond, Validation of a simple model accounting for light and temperature effect on microalgal growth, Bioresour Technol 123 (2012) 520–527.

[17] R. J. Geider, Light and temperature dependence of the carbon to chlorophyll a ratio in microalgae and cyanobacteria: implications for physiology and growth of phytoplankton, New Phytologist 106 (1987) 1–34.

[18] L. Rosso, J. R. Lobry, J. P. Flandrois, An unexpected correlation between cardinal temperatures of microbial growth highlighted by a new model., J Theor Biol 162 (4) (1993) 447–463.

[19] J. C. H. Peeters, P. Eilers, The relationship between light intensity and photosynthesis—A simple mathematical model, Hydrobiological Bulletin 12 (1978) 134–136.

[20] E. M. Bendif, I. Probert, D. C. Schroeder, C. de Vargas, On the description of *Tisochrysis lutea* gen. nov. sp. nov. and *Isochrysis nuda* sp. nov. in the Isochrysidales, and the transfer of *Dicrateria* to the Prymnesiales (Haptophyta), Journal of Applied Phycology .

[21] F. Mairet, O. Bernard, P. Masci, T. Lacour, A. Sciandra, Modelling neutral lipid production by the microalga *Isochrysis aff. galbana* under nitrogen limitation., Bioresour Technol 102 (2011) 142–149.

[22] S. M. Renaud, L.-V. Thinh, G. Lambrinidis, D. L. Parry, Effect of temperature on growth, chemical composition and fatty acid composition of tropical Australian microalgae grown in batch cultures, Aquaculture 211 (1) (2002) 195–214.

[23] R. Schenkendorf, A. Kremling, M.Mangold, Optimal experimental design with the sigma point method, IET Syst. Biol 3 (2009) 10–23.

[24] D. De Pauw, Optimal experimental design for calibration of bioprocess models: a validated software toolbox, Ph.D. thesis, University of Ghent, 2005.

[25] R. Muñoz-Tamayo, B. Laroche, M. Leclerc, E. Walter, IDEAS: a Parameter Identification Toolbox with Symbolic Analysis of Uncertainty and its Application to Biological Modelling, in: Prepints of the 15th IFAC Symposium on System Identification, Saint-Malo, France, 1271–1276, URL `http://www.inra.fr/miaj/public/logiciels/ideas/index.html`, 2009.

[26] T. Lacour, A. Sciandra, A. Talec, P. Mayzaud, O. Bernard, Diel variations of carbohydrates and neutral lipids in nitrogen-sufficient and nitrogen-starved cyclostat cultures of *Isochrysis* sp., Journal of Phycology 48 (2012) 966–975.

[27] F. Mairet, O. Bernard, T. Lacour, A. Sciandra, Modelling microalgae growth in nitrogen limited photobiorector for estimating biomass, carbohydrate and neutral lipid productivities, in: Proc. 18th World Congress The International Federation of Automatic Control, Milano, Italy, 10591–10596, 2011.

[28] E. Van Derlinden, K. Bernaerts, J. F. Van Impe, Accurate estimation of cardinal growth temperatures of *Escherichia coli* from optimal dynamic experiments., Int J Food Microbiol 128 (2008) 89–100.

[29] B. Chachuat, A. B. Singer, P. I. Barton, Global Methods for Dynamic Optimization and Mixed-Integer Dynamic Optimization, Ind. Eng. Chem. 45 (2006) 8373–8392.

[30] L. T. Biegler, Nonlinear Programming: Concepts, Algorithms, and Applications to Chemical Processes, Society for Industrial and Applied Mathematics and the Mathematical Optimization Society, Philadelphia„ 2010.

[31] J. Bonnans, Frédéric, P. Martinon, V. Grélard, Bocop - A collection of examples, Tech. Rep., INRIA, URL `http://hal.inria.fr/hal-00726992`, rR-8053, 2012.

[32] A. Wächter, T. Biegler, On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming, Mathematical Programming 106 (2006) 25–57.

[33] M. Rodríguez-Fernández, J. A. Egea, J. R. Banga, Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems., BMC Bioinformatics 7 (2006) 483.

[34] J. A. Egea, M. Rodríguez-Fernández, J. R. Banga, R., Martí, Scatter search for chemical and bio-process optimization, Journal of Global Optimization 37 (2007) 481–503.

[35] E. Van Derlinden, L. Mertens, J. F. Van Impe, The impact of experiment design on the parameter estimation of cardinal parameter models in predictive microbiology, Food Control 29 (2013) 300–308.

[36] D. Zou, K. Gao, Thermal acclimation of respiration and photosynthesis in the marine macroalga *Gracilaria lemaneiformis* (Gracilariales, Rhodophyta), Journal of Phycology .

[37] T. Turányi, Sensitivity analysis of complex kinetic systems. Tools and applications, Journal of Mathematical Chemistry 5 (3) (1990) 203–248.

[38] M. Ras, J.-P. Steyer, O. Bernard, Temperature effect on microalgae: a crucial factor for outdoor production, Reviews in Environmental Science and Biotechnology 12 (2013) 153–164.

[39] H. Pohjanpalo, System Identifiability Based on the Power Series Expansion of the Solution, Math Biosci 41 (1978) 21–33.

Figure 1: The TIP system. The device has 18 batch photobioreactors for microalgae cultivation.

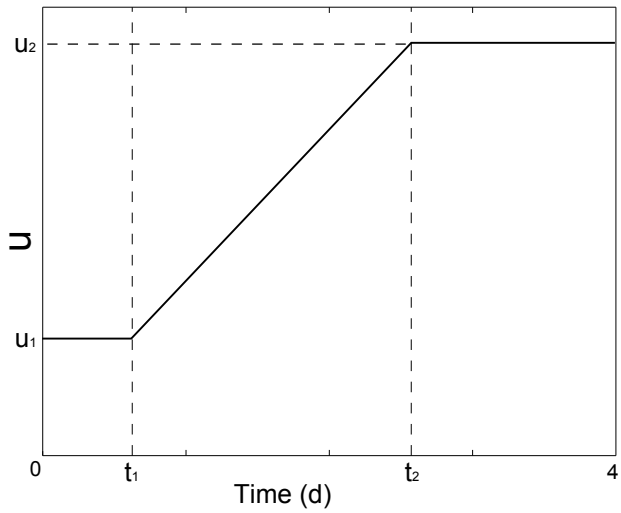

Figure 2: Parameterization of the experiment inputs u (T, I) for the CVP approach.

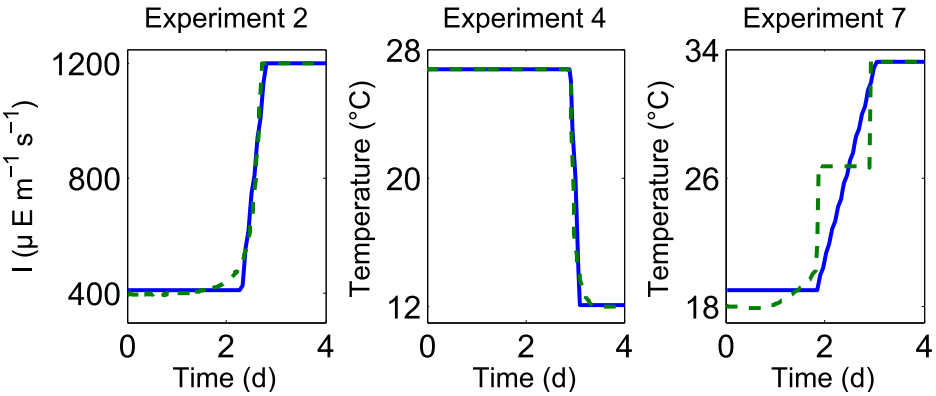

Figure 3: Optimal experiment inputs given by the CVP approach (solid lines) and the simultaneous approach (dashed lines) for the partitioned OED problem.

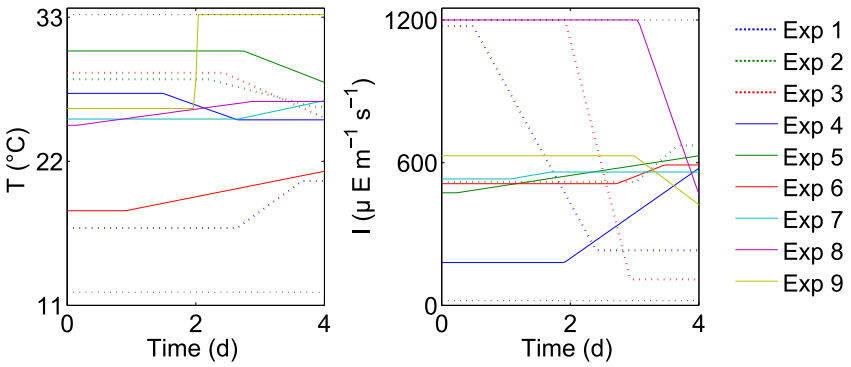

Figure 4: Optimal experiment inputs obtained for the full OED problem.

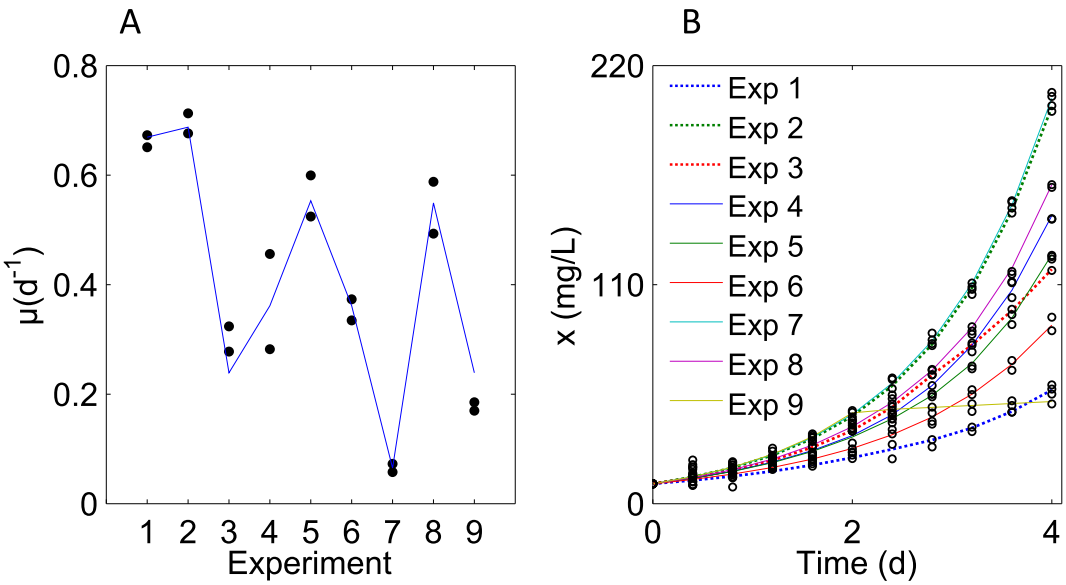

Figure 5: Simulated data resulted from D-optimal experiments including duplicates and responses of the identified models. A. Measurements of specific growth (circles) for the static approach. B. Measurements of biomass concentrations (circles) for the dynamic approach obtained for the full OED problem. The responses of the identified models (solid and dotted lines) for both static and dynamic approaches described satisfactorily the simulated data.

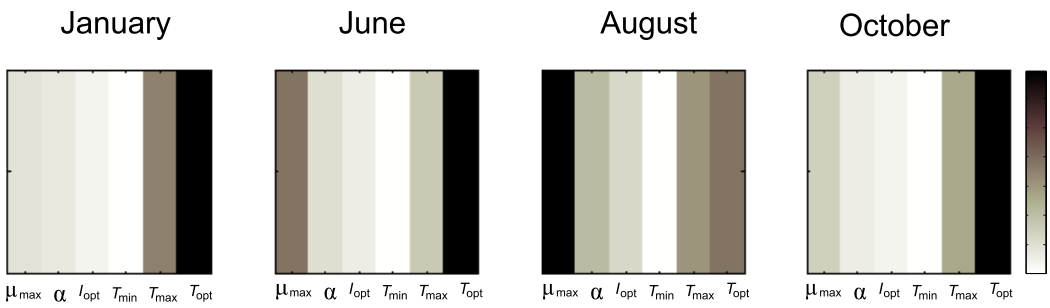

Figure 6: Overall sensitivity of the biomass concentration to the parameters in the complete raceway model developed in [3]. Results are shown for four months illustrating how the influence of the parameters on the model output is modulated by the environmental conditions.

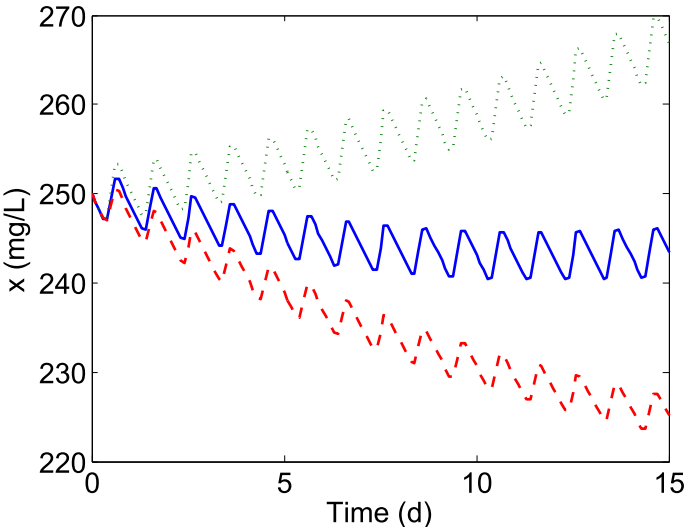

Figure 7: A small uncertainty of 5% on the value of T_{opt} leads to important mismatches on model predictions. The dynamic of the biomass concentration of the month of January with the nominal value of T_{opt} (solid blue line) is compared to the response of the model with $0.95T_{\text{opt}}$ (dotted green line) and $1.05T_{\text{opt}}$ (dashed red line) .

Table 1: Nominal values of the model parameters.

Parameter	Definition	Value
α	Initial slope of the growth response curve w.r.t. light	0.008 (μE m^{-2}s^{-1} d)$^{-1}$
$\mu_{\max}$	Maximal specific growth rate	0.76 d^{-1}
I_{opt}	Optimal light intensity	548 μE m^{-2}s^{-1}
$T_{\min}$	Lower temperature for microalgae growth	-0.20 °C
$T_{\max}$	Upper temperature for microalgae growth	33.30 °C
T_{opt}	Temperature at which growth rate is maximal	26.70 °C

Table 2: Subproblems of the dynamic OED strategy.

Experiment	Couple of parameters	Experiment input
1	$(\mu_{\max}, \alpha)$	I
2	$(\mu_{\max}, I_{\text{opt}})$	I
3	(α, I_{opt})	I
4	$(\mu_{\max}, T_{\min})$	T
5	$(\mu_{\max}, T_{\max})$	T
6	$(\mu_{\max}, T_{\text{opt}})$	T
7	$(T_{\min}, T_{\max})$	T
8	$(T_{\min}, T_{\text{opt}})$	T
9	$(T_{\max}, T_{\text{opt}})$	T

Table 3: Experimental conditions for the static approach.

Full factorial design									
	Experiment								
	1	2	3	4	5	6	7	8	9
Temperature (°C)	12.0	22.60	33.20	12.0	22.60	33.20	12.0	22.60	33.20
Light intensity (μE $m^{-2}s^{-1}$)	20	20	20	610	610	610	1200	1200	1200
Composite factorial design								[14]	
	Experiment								
	1	2	3	4	5	6	7	8	9
Temperature (°C)	15.30	19.0	19.0	19.0	19.0	24.5	24.5	24.5	24.5
Light intensity (μE $m^{-2}s^{-1}$)	560	863	257	257	863	560	560	560	560
	10	11	12	13	14	15	16	17	
Temperature (°C)	24.50	24.50	24.50	30.0	30.0	30.0	30.0	33.20	
Light intensity (μE $m^{-2}s^{-1}$)	560	50	1070	863	257	257	863	560	
D-optimal design									
	Experiment								
	1	2	3	4	5	6	7	8	9
Temperature (°C)	12.10	12.10	24.30	24.60	26.70	26.70	30.60	30.70	33.20
Light intensity (μE $m^{-2}s^{-1}$)	536	536	1200	409	74	74	1200	395	547

Table 4: Comparison of D-optimal design with factorial design for the static approach.

	$\mu_{\max}$	α	I_{opt}	$T_{\min}$	$T_{\max}$	T_{opt}
$\frac{\eta_j \text{D-optimal}}{\eta_j \text{Composite factorial design}}$	0.50	0.48	0.76	0.16	0.02	0.25
	$\text{Det}(\mathbf{F}_{\text{s}})$			$\lambda_{\max}/\lambda_{\min}$		
Full factorial design	0			$5.30{\cdot}10^{20}$		
Composite factorial design	381.74			$1.89{\cdot}10^{9}$		
D-optimal	$7.90 \cdot 10^{6}$			$3.85 \cdot 10^{9}$		

Table 5: Comparison of the CVP and sequential strategies for the partitioned OED problem in the dynamic approach.

Experiment	$-\log\mathrm{Det}\left(\mathbf{F}_{\mathrm{d}}\right)$		
	Simultaneous (J_{Sim})	CVP (J_{cvp})	$J_{\mathrm{cvp}}/J_{\mathrm{Sim}}$
1	-26.1949	-25.8114	0.9854
2	-5.37552	-5.0026	0.9306
3	-10.2537	-9.8398	0.9596
4	-11.3121	-10.8970	0.9633
5	-16.9941	-16.5543	0.9741
6	-14.856	-14.6943	0.9891
7	-5.0363	-4.4793	0.8894
8	-4.5046	-4.0965	0.9094
9	-9.9769	-9.4997	0.9522

Table 6: Comparison of the accuracy of the estimation obtained with the solutions of the full and partitioned OED problems in the dynamic approach.

	$\mu_{\max}$	α	I_{opt}	$T_{\min}$	$T_{\max}$	T_{opt}
$\frac{\eta_j \text{Full}}{\eta_j \text{Partitioned}}$	0.85	0.57	0.41	0.76	1.05	0.59
	$\text{Det}(\mathbf{F}_{\text{d}})$			$\lambda_{\max}/\lambda_{\min}$		
Full OED	$1.76{\cdot}10^{12}$			$1.10{\cdot}10^{9}$		
Partitioned OED	$4.20{\cdot}10^{10}$			$4.44{\cdot}10^{9}$		

Table 7: Estimated parameters with their approximate confidence intervals for the static and dynamic OED approaches. The parameter estimation was performed with simulated noisy data.

	$\hat{\theta} \pm 2\eta_j$					
	$\mu_{\max}$	α	I_{opt}	$T_{\min}$	$T_{\max}$	T_{opt}
Static	0.74±0.070	0.0075±0.0029	665.03±274.52	-0.86±4.84	33.34±0.30	26.80±0.97
Dynamic	0.76±0.0091	0.008±0.0014	550.64±35.44	-0.26±2.80	33.34±0.28	26.66±0.25

Appendix A. Comments on structural and practical identifiability of the model

Appendix A.1. Structural identifiability

To check the structural identifiability of the model, we used the time power series method [39], briefly described below.

Consider the following model

$$\dot{x}(t) = f(x(t), u(t), \boldsymbol{\theta}, t), \tag{A.1}$$

$$y_{\mathrm{m}}(t) = h(x(t), \boldsymbol{\theta}), \tag{A.2}$$

with x the state vector, y_{m} the vector of model outputs(observations) and $u(t)$ the control vector. The function vector $f(\cdot)$ is assumed to have infinitely many derivatives with respect to time and the input and state vector components. In the same manner, $h(\cdot)$ is infinitely differentiable w.r.t. the state vector and $u(t)$ is infinitely differentiable w.r.t. time. Both the state and the outputs are infinitely differentiable w.r.t. time. The outputs can therefore be represented by the corresponding Taylor series expansion around $t = 0$.

Consider the kth time derivative (a_k) of the output.

$$a_k(\boldsymbol{\theta}) = \lim_{t \to 0+} \frac{\mathrm{d}^k}{\mathrm{d}t^k} y_{\mathrm{m}}(t). \tag{A.3}$$

Since all the time derivatives of the outputs are unique, it follows that a sufficient condition for the identifiability of the model is that set of equations

$$h^{(k)}(x(0), \boldsymbol{\theta}) = a_k(0) \tag{A.4}$$

have a unique solution for $\boldsymbol{\theta}$ [39].

For our case study, let us consider first the identifiability of the temperature parameters of the cardinal model $T_{\mathrm{min}}, T_{\mathrm{max}}, T_{\mathrm{opt}}$. At constant light, the model is given by

$$\dot{x}(t) = \mu_I \phi_T(t) x(t), \quad x(0) = x_0, \tag{A.5}$$

$$y_{\mathrm{m}}(t) = x(t), \tag{A.6}$$

with $\mu_I = \mu_{\mathrm{max}} \phi_I$ and x_0 a known initial concentration of biomass. The effects of light ϕ_I and temperature ϕ_T on microalgae growth are here recalled:

$$\phi_I = \frac{I}{I + \frac{\mu_{\max}}{\alpha}\left(\frac{I}{I_{\text{opt}}} - 1\right)^2}, \tag{A.7}$$

553

$$\phi_T = \begin{cases} 0, & T < T_{\min} \\ \frac{(T-T_{\max})(T-T_{\min})^2}{(T_{\text{opt}}-T_{\min})[(T_{\text{opt}}-T_{\min})(T-T_{\text{opt}})-(T_{\text{opt}}-T_{\max})(T_{\text{opt}}+T_{\min}-2T)]}, & T \in [T_{\min}, T_{\max}] \\ 0, & T > T_{\max}. \end{cases} \tag{A.8}$$

554 By simple inspection of (A.8) and given the biological meaning of the parameters
555 of the cardinal model, we can infer that a series of adequate experiments running at
556 different temperature conditions in the interval $[T_{\min}, T_{\max}]$ will allow to identify uniquely
557 the temperature parameters.

558 Coming back to the time power series method, let us consider the case of a specific
559 input $T(t)$ that is infinitely differentiable w.r.t. to time. For simplicity, we chose $T(t) =$
560 $c_1 t + c_2$ with $c_1 > 0, c_2 > 0$ and $T(t) \in [T_{\min}, T_{\max}]$ in the experimentation time.

The Taylor series coefficients are thus

$$a_0 = x_0, \tag{A.9}$$

$$a_1 = \mu_I \phi_T(0) x_0, \tag{A.10}$$

$$a_2 = \left((\mu_I \phi_T)^2 + \mu_I \left[\frac{\partial \phi_T}{\partial T} \right]_{T=T(0)} \dot{T}(0) \right) x_0. \tag{A.11}$$

Given the shaphe of ϕ_T and applying the first-optimality condition, the following cases provide the parameters to be uniquely identifiable:

$$a_1 = 0, \text{and}, \ a_2 \geq 0, \ \Rightarrow T_{\min} = T(0), \tag{A.12}$$

$$a_1 = 0, \text{and}, \ a_2 < 0, \ \Rightarrow T_{\max} = T(0), \tag{A.13}$$

$$a_1 > 0, \text{and}, \ a_2 = a_1^2/x_0, \ \Rightarrow T_{\text{opt}} = T(0). \tag{A.14}$$

561 The previous conditions can be reached by making $T(t)$ vary along the interval $[T_{\min}, T_{\max}]$.

Following the same procedure, we can now check the identifiability of the light parameters I_{opt}, α and the maximal specific growth rate $\mu_{\max}$. Since T_{opt} is structurally identifiable, let us consider a constant temperature $T = T_{\text{opt}}$. The model is thus

$$\dot{x}(t) = \mu_{\max} \phi_I(t) x(t), \quad x(0) = x_0, \tag{A.15}$$

$$y_{\text{m}}(t) = x(t). \tag{A.16}$$

By considering the case of a specific input $I(t) = c_1 t + c_2$, the series expansion provides the following coefficients

$$a_0 = x_0, \tag{A.17}$$

$$a_1 = \mu_{\max}\phi_I(0)x_0, \tag{A.18}$$

$$a_2 = \left((\mu_{\max}\phi_I(0))^2 + \mu_{\max}\left[\frac{\partial\phi_I}{\partial I}\right]_{I=I(0)}\dot{I}(0)\right)x_0. \tag{A.19}$$

By applying the first-order optimality condition on ϕ_I, we get

$$a_1 > 0, \text{and } a_2 = a_1^2/x_0, \Rightarrow I_{\text{opt}} = I(0). \tag{A.20}$$

Injecting I_{opt} in (A.18) provides $\mu_{\max}$.

Once I_{opt} is identified, evaluating (A.18) at $I(0) \neq I_{\text{opt}}$ provides α

$$\alpha = \frac{a_1\mu_{\max}}{I(0)(\mu_{\max}x_0 - a1)}\left(\frac{I(0)}{I_{\text{opt}}} - 1\right)^2. \tag{A.21}$$

The model is therefore structurally identifiable.

Appendix A.2. Practical identifiability

Parameter estimation of Haldane and Monod type kinetics is known to be hampered by practical identifiability problems due to strong correlation between its parameters. To represent the effect of light on microalgae growth, the Haldane kinetics is often used

$$\phi_I = \tilde{\mu}\frac{I}{I + K_{sI} + I^2/K_{iI}}, \tag{A.22}$$

where $\tilde{\mu}$ is the specific growth rate, K_{sI} is the light affinity constant and K_{iI} is the inhibition constant. By applying the first-order optimality condition, the optimal light intensity for growth is $I_{\text{opt}} = \sqrt{K_{sI}K_{iI}}$. The nominal values for the Haldane model used in the present study were $\tilde{\mu} = 1.18$ d^{-1}, $K_{sI} = 150$ μE m^{-2}s^{-1} and $K_{sI} = 2000$ μE m^{-2}s^{-1}.

In this work, instead of the standard Haldane kinetics, we made use of the parameterized kinetics (A.7), which has the same shape than the Haldane kinetics but offers certain advantages in terms of practical identifiability properties. On the basis of good quality nominal parameters, it was previously shown that adequate inputs allow to identify the optimal conditions for growth I_{opt} and T_{opt}, which derive automatically on the

identification of $\mu_{\max}$. Such property allows to weaken the correlation between $\mu_{\max}$ and α.

For illustration, we performed a D-optimal protocol consisted of five experiments with ten equidistant sampling times for both standard Haldane and the parameterized kinetics.

For the Haldane kinetics, the correlation matrix of the parameters was:

$$\begin{array}{cccc} \tilde{\mu} & 1.0 & & \\ K_{sI} & 0.96 & 1.0 & \\ K_{iI} & -0.98 & -0.92 & 1.0 \end{array}$$

For the parameterized kinetics, the correlation matrix of the parameters was:

$$\begin{array}{cccc} \mu_{\max} & 1.0 & & \\ \alpha & -0.53 & 1.0 & \\ I_{\text{opt}} & -0.25 & -0.10 & 1.0 \end{array}$$

As observed, the Haldane kinetics exhibits a stronger parameter correlation than the parameterized kinetics. Therefore, in terms of practical identifiability, the parameterized kinetics is preferred over the standard Haldane kinetics.